CHEMICAL RESEARCH, DEVELOPMENT & ENGINEERING CENTER

CRDEC-TR-085

USING THEORETICAL DESCRIPTORS
IN STRUCTURAL ACTIVITY RELATIONSHIPS
V. A REVIEW OF THE THEORETICAL PARAMETERS

George R. Famini, Ph.D.

RESEARCH DIRECTORATE

July 1989

AD-A213 580

DTIC
ELECTE
OCT 2 4 1989
S D
D^{c_3}

U.S. ARMY
ARMAMENT
MUNITIONS
CHEMICAL COMMAND

Aberdeen Proving Ground, Maryland 21010-5423

89 10 23 090

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

Distribution Statement

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRDEC-TR-085	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION CRDEC	6b. OFFICE SYMBOL (If applicable) SMCCR-RSP-C	7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21010-5423	7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION CRDEC	8b. OFFICE SYMBOL (If applicable) SMCCR-RSP-C	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)

Aberdeen Proving Ground, MD 21010-5423

10. SOURCE OF FUNDING NUMBERS			
PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
	1C162622	A5531	

11. TITLE (Include Security Classification)
Using Theoretical Descriptors in Structural Activity Relationships
V. A Review of the Theoretical Parameters

12. PERSONAL AUTHOR(S)
Famini, George R., Ph.D.

13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM 88 Jan TO 88 Jul	14. DATE OF REPORT (Year, Month, Day) 1989 July	15. PAGE COUNT 25

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	QSAR Structural activity relationships,
12	04		LSER ORGAN...
15	06	03	MOPAC

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Quantitative structural activity relationships (QSAR) have been used successfully in the past to develop predictive equations for numerous biological and physicochemical properties. Kamlet and Taft use linear solvation energy relationships (LSER), a subset of QSAR, to correlate over 100 solute/solvent dependent interactions with a set of empirically derived descriptors. A major difficulty with this approach is the use of these empirically determined parameters. A new approach based upon the generalized LSER approach has been developed. This technique, called theoretical LSER, uses only structural and quantum chemically derived descriptors. A review of the parameter set and a comparison of correlations to the LSER are described.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL SANDRA J. JOHNSON	22b. TELEPHONE (Include Area Code) (301) 671-2914	22c. OFFICE SYMBOL SMCCR-SPS-T

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

PREFACE

The work described in this report was authorized under Project No. 1C162622A553I, CB Simulants, Survivability, and Systems Science. This work was started in January 1988 and completed in July 1988.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

Reproduction of this document in whole or in part is prohibited except with permission of the Commander, U.S. Army Chemical Research, Development and Engineering Center, ATTN: SMCCR-SPS-T, Aberdeen Proving Ground, Maryland 21010-5423. However, the Defense Technical Information Center and the National Technical Information Service are authorized to reproduce the document for U.S. Government purposes.

This report has been approved for release to the public.

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and/or Special	
A-1		

3

Blank

4

CONTENTS

LIST OF FIGURES

LIST OF TABLES

USING THEORETICAL DESCRIPTORS IN STRUCTURAL ACTIVITY RELATIONSHIPS
V. A REVIEW OF THE THEORETICAL PARAMETERS

1. INTRODUCTION

The estimation and prediction of physical and chemical properties of compounds and associated activities generated a great deal of interest in the past several years. A principal reason for this has been the need to know various properties quickly and often before the compound has been synthesized. There are two excellent compilations of property estimation routines in the literature.[1,2]

An effort is underway at the U.S. Army Chemical Research, Development and Engineering Center (CRDEC) to develop the capability for a priori prediction of physical and chemical properties.[3] Although many of the routines in the above compilations are useful, the unique nature of the compounds of interest often requires new methods to be developed or existing methods to be modified.

1.1 Linear Free Energy Relationships.

Although it has been theorized for quite some time that molecular structure directly impacts physical and chemical activities, it was not until Hammett developed the concept of Linear Free Energy Relationships (LFER) that the relationship was quantified.[4] Using heuristic thermodynamic and kinetic arguments, Hammett developed a theoretically supportable equation that related a given reactivity to substituent effects.

The basic premise behind LFER is shown in equation 1 where the equilibrium constant (K_{eq}) is related to the Gibbs Free Energy (ΔG).

$$\Delta G = -2.30 \; RT \; LOG \; K_{eq} \tag{1}$$

Hammett examined, in particular, the hydrolysis of benzoic acid ester and the effects of substitution on the relative rates of reaction. From simple substitution of equation 1, he was able to derive equation 2.

$$LOG \; \frac{K}{K_0} = \rho\sigma \tag{2}$$

The terms ρ and σ represent electronic effects due to the nature of the substituents on the benzene (σ) or to the nature of the reaction (ρ).

1.2 Quantitative Structural Activity Relationships.

Using the arguments of Hammett, Hansch expanded the concept of the LFER to increase the use of this type of equation.[5] Two assumptions were made: the activity being considered was directly related to the ΔG, and the ΔG was directly related to the molecular description.[6] This relationship is shown in equation 3.

$$LOG \; A = \Sigma a_i x_i \tag{3}$$

where x_i represents the ith molecular descriptor/parameter in the derived regression, and a_i represents the corresponding coefficients. A wide variety of parameters have been used in the past to achieve the most consistent correlations.

1.3 Linear Solvation Energy Relationships.

Using the procedures of Quantitative Structure Activity Relationship (QSAR) and LFER, Kamlet and Taft developed a relationship correlating a specialized set of parameters with a large number of solute/solvent interactions.[7-9] The relationship, the Generalized Linear Solvation Energy Relationship (LSER), takes the form shown in equation 4.

$$\text{Log Property} = \text{Steric Term} + \text{Polarizability Term} + \text{Hydrogen Bonding Term(s)} + C \tag{4}$$

Kamlet and Taft found three principal factors affecting a given solute/solvent interaction: a. steric effects, the size constraints of the solute and the ability of the solvent to accommodate the particular size of the solute; b. polarizability/dipolarity, the ability to have a dipole induced in a molecule; and c. the ability to initiate or accept hydrogen bonding. In following the reasonings of QSAR and LFER, a series of parameters were developed that reflect each of these general factors. The parameters, often called the Kamlet-Taft or solvatochromic parameters, were correlated with a large number of different types of solute/solvent interactions.[9]

The LSER equation with the appropriate solvatochromic parameters is shown in equation 5. These descriptors are explained in Table 1.

$$\text{LOG Property} = mV_m + p\pi^* + a_\alpha + b_\beta + C \tag{5}$$

1.4 The Theoretical LSER.

The descriptors in Table 1 were a tremendous aid in advancing the understanding of solute/solvent interactions from a physical-organic standpoint. One major difficulty in this development is the nature of the solvatochromic parameters. Each of these parameters are inherently empirical in nature, preventing the use of these equations for a prior prediction and estimation of solute/solvent properties.

Theoretically derived parameters are now being used in these types of relationships.[10] The previous reports in this series have dealt with development of the Theorectical LSER (TLSER) and the applications to the prediction and estimation of physical and chemical properties.[11-14] This report is an attempt to summarize the TLSER and to demonstrate the usefulness of this procedure. The generalized TLSER is shown in equation 6.

$$\text{LOG Property} = mV_{mc} + p\pi_I + a_1\epsilon_a + a_2qH_+ + b_1\epsilon_b + b_2q_- + C \tag{6}$$

The parameters are shown in Table 1 and fully described in the Results section.

Table 1. Symbols Used

MW	=	Molecular Weight (g/mol)
d_I	=	Liquid density (g/ml)
V_m	=	Molar Volume
π^*	=	Solvatochromic Polarizability Term
β	=	Solvatochromic Hydrogen Bond Basicity Term
α	=	Solvatochromic Hydrogen Bond Acidity Term
V_{mc}	=	Molecular Volume (A^3)
π_I	=	Polarizability Index
ϵ_b	=	Molecular Orbital Basicity (kcals)
ϵ_a	=	Molecular Orbital Acidity (kcals)
q_-	=	Electrostatic Basicity
qH_+	=	Electrostatic Acidity
K_{ow}	=	Octanol/Water Partition Coefficient
EA	=	Electronic Absorption (UV-Vis)
LC_{50}	=	50% Lethal Concentration (Konneman)
EC_{50}	=	50% Effective Concentration (Microtox)
k_{75}	=	HPLC Retention Capacity in 75% Methanol
k_{OH}	=	Hydrolysis Rate Constant
C	=	Regression Coefficient

2. EXPERIMENTATION

2.1 Experimental Data.

All experimental data were taken either from the original sources for the data or from the corresponding report by Kamlet and Taft. To maintain as much consistency as possible, where the TLSER and LSER are directly compared, identical data sets have been used. The α, β, π^* and molar volumes were taken from articles of Kamlet and Taft.[8]

2.2 Calculations.

All calculations, regressions, (results of the regressions and data appear in the previous reports of this series). All calculations were performed on a μVAXII running under the VMS Operating Systems. All quantum calculations and geometric optimizations were performed using the MNDO algorithm within the MOPAC.[15,16] The molecualr volumes were calculated using the procedure of Hopfinger incorporated in the MMADS software.[17,18] All regressions were performed uning the MINITAB statistical software package.*

3. RESULTS

The development of the TLSER can be envisioned as two separate goals. First, the developed TLSER parameters should, as much as possible, correlate

---

*MINITAB is marketed by Minitab, INC., State College, PA

9

one for one with the LSER parameters. This will permit the estimation of the LSER parameters. There have been some attempts at this although the correlations obtained are of questionable value.[10] This could be due to the level of the theoretical calculations as well as the parameters used. The second goal is the direct implementation of the theoretical descriptors into the TLSER equation, bypassing the LSER parameters. In fact, this goal is the more important, as the primary goal is the development of usable methods of a priori prediction of physical and chemical properties. There have been quite a few correlations of theoretical parameters in QSAR equations although, again, the correlations have usually been less than satisfactory for predictive purposes.[19] The following sections describe the TLSER parameters, compare them to the LSER parameters, then combine all of them into direct correlations with the solute/solvent properties.

Table 1 lists the abbreviations used to refer to the LSER and TLSER parameters throughout this report. All equations are reported with the data set size, N, and two descriptions of the degree of fit: the standard error of the estimate, SEE (sd); and the correlation coefficient, R.

3.1   The TLSER Descriptors.

3.1.1   Molecular Volume.

The first term in the LSER equation is a steric or cavity term. This descriptor is intended as a measure of the energy required to create a hole in the solvent matrix sufficient to solvate the solute molecule. In dealing with multiple solutes in a single solvent, the energy required to separate the solvent molecules a given distance remains constant, but the distance required to separate the molecules varies with the size of the solute molecule. In the LSER, the steric term used to account for this phenomenon is the Molar Volume (V_m, MW/d_i).

In the past, steric terms have been the easiest to calculate, because electronic contributions are not explicitly considered in this descriptor. There have been many steric terms used in QSAR equations with very good results, ranging from such empirically derived descriptors as Molar Volume and Taft's E_s parameter to theoretically derived volumes and surface areas.[5] Using computational techniques, it is possible to calculate the volume taken up by a molecule in space by adding the Van der Waal's contributions for each atom and subtracting out the overlap volumes. Figure 1 shows a pictorial representation of the Molecular Volume of a methyl chloride molecule.
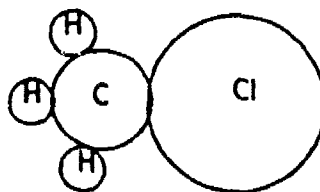

Figure 1. Molecular Volume

In theory, then, the V_{mc} and the V_m should correlate well with each other. Equation 7 shows the relationship between these two terms for 80 diverse compounds.[11]

$$\frac{V_m}{100} = 0.891 \left[\frac{V_{mc}}{100} \right] + 0.152$$

(7)

$$N = 80 \quad R = 0.9690 \quad SEE = 0.073$$

From equation 7, it can be seen that the fit with the molar volume is excellent, indicating that the V_{mc} can be used to predict and to replace the V_m.

3.1.2   Polarizability Index.

The remaining terms in the LSER, and therefore the TLSER, are electronic in nature. To best approximate these methods, quantum chemical methods were employed. Several methods were compared for computational speed and the ability to generate reasonable geometries, including Gaussian82, ZINDO, MNDO, and AMPAC. The MNDO method with MOPAC was chosen as giving the most reasonable geometry for the data sets employed without requiring an unreasonable amount of computer time.

The polarizability (π^*) was the second term of the LSER to be replaced. In general, this term describes the ability of the electrons in the valence shell of the solute to be polarized. This is shown in Figure 2, where the electron cloud has a roughly elliptical shape in the absence of any field. As an electric field is introduced, the electrons polarize themselves to cause a dipole moment to be induced in the molecule. In normal solution chemistry, this field is caused by solvents or other compounds with high dipole moments.
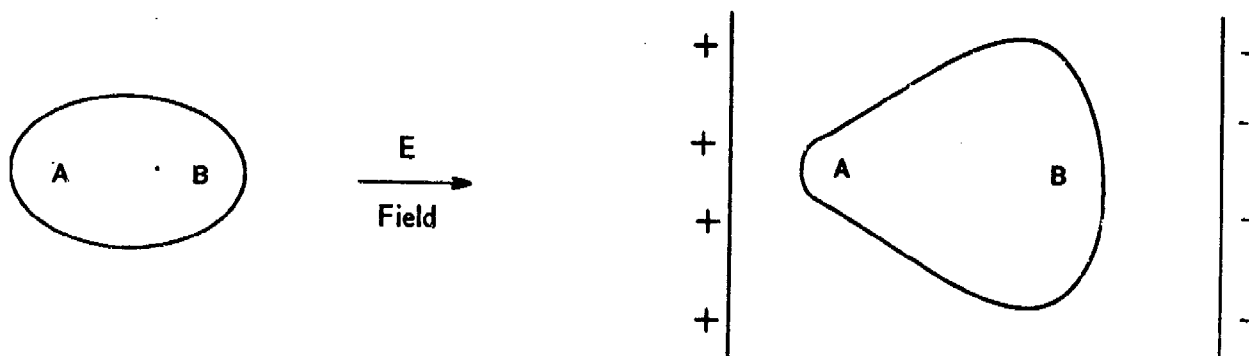

Figure 2.   Polarizability

MOPAC is capable of calculating the Polarization Volume (PV) using the output of a MNDO calculation.[20] In effect, the PV is the change in the Van der Waal's volume of the molecule due to the introduction of an electrical field. This term, with units of A^3, has a high correlation with the Molecular

11

Volume and, in effect, is a combination of volume and to generate a unitless descriptor, the Polarizability Index (π_I) is calculated as shown in equation 8.[12]

$$\pi_I = \frac{\text{Polarization Volume}}{\text{Molecular Volume}} \tag{8}$$

This results is a descriptor which is solely dependent upon the ability of the electrons to move throughout the entire molecule. As such, one would expect aromatic compounds to be among the highest values with double and triple bonds to be below that. Compounds with highly electronegative atoms would be expected to have lower π_I values, and those with third and fourth row atoms greater. Table 2 shows five representative π, values. showing this comparison.

Table 2. Representative Polarizability Index Values

Compound	π_I
Ethanol	0.0927
Propane	0.0953
Butyl Chloride	0.1036
Benzene	0.1204
Pentachlorobenzene	0.1331

3.1.3   Basicity Terms.

Two terms were required to adequately correlate with the β term of the LSER.[14] These two terms are the Molecular Orbital Basicity (ϵ_b) and the Electrostatic Basicity (q_-). Equation 9 shows the correlation with β.

$$\beta = 0.240 - 0.196\epsilon_b + 1.699q_-$$

$$N = 77 \quad R = .9518 \quad SEE = 0.0777 \tag{9}$$

3.1.4   Molecular Orbital Basicity.

The Molecular Orbital Basicity can be envisioned both in theoretical and chemical terms. Equation 10 shows the relationship for this term to the Highest Occupied Molecular Orbital (HOMO) and the Lowest Occupied Molecular Orbital (LUMO).

$$\epsilon_b = E_{LUMO_{Water}} - E_{HOMO_{Substrate}} \tag{10}$$

E_{LUMO} is the energy of the LUMO of water, and E_{HOMO} is the HOMO of the substrate or solute. Therefore, the ϵ_b is the interaction energy for the covalent transfer of electrons from the substrate to water. The lower the ϵ_b, the more facile the exchange. This term is a covalent basicity term and is shown in Figure 3. Table 3 lists representative values for the β, ϵ_b, and qH.
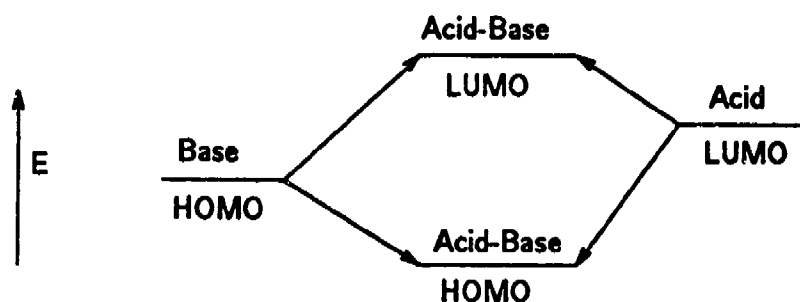
Figure 3. Molecular Orbital Basicity

Table 3. Representative Values of Basicity

Compound	ϵ_b	q_-	β
Carbon Tetrachloride	18.6798	0.070	0.10
Hexane	17.4761		0.022
Butanol	16.7365	0.325	0.45
Benzene	14.8338	0.059	0.10
Hexamethyl Phosphoramide	15.4321	0.650	1.05

According to Ligand Field Theory, when an acid-base interaction (or any reaction) occurs, the HOMO of base and LUMO of the acid mix to form an adduct HOMO and an adduct LUMO.[20] The nearer the HOMO and LUMO of the reactants are in energy, the greater the degree of overlap results. The greater degree of overlap lowers the adduct HOMO. In general, the lower adduct HOMO signifies a more stable product and, therefore, a more likely reaction.

This can be related to the TLSER parameters and to the LSER β. Because ϵ_b represents the difference between the HOMO of the substrate and the LUMO of water, the larger the difference, the lower the basicity. This implies that an acid-base interation would be less likely to occur. Therefore ϵ_b should have a negative sign in the correlation for β and an opposite other than that of β in the property correlations. Going back to equation 9, it can be seen that β requires a negative sign. The validity of an opposite other than β in the property correlations shall be shown in later sections. Table 3 lists a representative set of ϵ_b and β values of five compounds.

3.1.5 Electrostatic Basicity.

The second basicity term, and the final parameter to be discussed in depth in this study, is the Electrostatic Basicity (q_-). This parameter is the absolute value of the most negative formal charge in the substrate. Where ϵ_b represents the covalent interaction of acid base, q_- represents the charge-charge attraction. It would be expected, then, as q_- gets larger

13

in absolute magnitude, the attraction would be better, and the substrate would be a better base. The β should then vary proportionally to q_-. This can be seen in equation 9 where the appropriate coefficient has a positive sign.

3.1.6 Acidity Terms.

In replacing the acidity term, α, in the LSER, an identical algorithm was employed for the basicity terms. Two parameters were developed, the Molecular Orbital Acidity (ϵ_a) and the Electrostatic Acidity (qH_+). A correlation has not yet been made between α with ϵ_a and qH_+. Table 4 lists corresponding ϵ_a, qH_+, and α values for five representative compounds.

Table 4. Representative Values of Acidity

Compound	ϵ_a	qH_+	α
Methanol	15.9831	0.236	0.33
Acetone	12.850	0.023	0.00
Benzene	12.5588	0.059	0.00
Phenol	12.4382	0.193	0.61
Carbon Tetrachloride	15.2253	0.000	0.00

3.1.7 Molecular Orbital Acidity.

The Molecular Orbital Acidity, ϵ_a, is analogous to ϵ_b. Rather than describing the flow of electrons from the substrate to water, it describes the flow of electrons from water to the substrate. The ϵ_a is merely the E_{HOMO} of water subtracted from the E_{LUMO} of the substrate. This is shown in equation 11.

$$\epsilon_a = E_{LUMO_{Substrate}} - E_{HOMO_{Water}} \tag{11}$$

As can been seen, this is the exact opposite of ϵ_b, with the substrate now acting as the acid and water acting as the base.

3.1.8 Electrostatic Acidity.

The Electrostatic Acidity, qH_+, represents the hydrogen or Bronstead acidity in the molecule. This is described by the most positive formal charge of any hydrogen atom in the substrate molecule.

3.2 Examples of the TLSER.

The real use of the TLSER parameters is not so much in being able to theoretically validate and predict the LSER parameters but rather to be able to be used in relationships with actual solute/solvent interactions. To provide a degree of comparison, corresponding LSER correlations will also be given with the TLSER correlations.

14

Several different solute/solvent interactions were examined, including toxicological properties, physical properties, and spectral absorptions. The following sections describe each interaction in depth.

3.2.1   Octanol/Water Partition Coefficient.

The Octanol/Water Partition Coefficient (K_{ow}) has been used extensively in medicinal chemistry and in QSAR equations.[22,23] The principal reason for this has been the degree to which the K_{ow} models the transport of compounds over the blood/brain barrier and other lipophilic/hydrophilic partitionings. As such, K_{ow} has become the principal parameter in QSAR equations involving biological activities.[19]

Using the TLSER parameters, equation 12 was developed for a data set of 64 compounds of diverse classes.[7,14]

$$LOG\ K_{ow} = 2.995\left[\frac{V_{mc}}{100}\right] - 0.847\ (\pi^*10) + 1.73\left[\frac{\epsilon b}{10}\right] - 5.415q_- - 3.960$$

$$N = 64 \quad R = 0.9566 \quad SEE = 0.3568$$

(12)

A similar relationship was developed by Kamlet and Taft using the LSER parameters shown in equation 13. It is of note that Kamlet and Taft included only general aliphatics and alcohols but did not included aromatics in the equation below.

$$LOG\ K_{ow} = 0.10 + 2.89\left[\frac{V_m}{100}\right] - 0.88\pi^* - 3.62\beta$$

$$N = 63 \quad R = .9890 \quad SEE = 0.1800$$

(13)

In comparing these two equations, several items are evident. First, the correlation coefficient (R) is significantly higher in the LSER equation (13) than in the TLSER equation (12), and the standard error of the estimate (SEE) is much lower. Two factors can account for this increased fit. First, the TLSER was developed for a gas-phase molecule in a vacuum with no intermolecular interactions taken into account. This is a major shortcoming of quantum chemical techniques in general. The LSER parameters were developed in solution, and take into account those solute factors the TLSER parameters cannot. At present, the exact magnitude these errors represent is not evident. The second factor relates to the use of empirical correction terms. Kamlet and Taft recognized that certain classes of compounds were routinely found as outliners in their LSER relationships. As such, they developed an elaborate set of correction terms to bring those faulty values into line. A correlation without these correction factors would yield a result much worse than the TLSERs. On the other hand, the TLSER has not had any corrections applied to any of the descriptors and, nevertheless, results in very high correlations.

With respect to the descriptors, there are several additional trends that are evident. First, the molecular volume and molar volume contain the same information. This can be seen because the coefficients are the same.

15

These are the most important factors in describing the realationship. This can be seen from Table 5, which lists the t-scores for each of the variables in Equation 12. In identification, the sign of the coefficient is important because larger compounds prefer the octanol layer, giving rise to a larger K_{ow} result.

Table 5.   t-Scores for Equation 12

Parameter	t-Score
V_{mc}	11.73
πI	-0.37
εb	2.76
q_-	-11.63

The second descriptor in each equation, π^* and $\pi 1$, also contains obvious similarities. Although the two parameters do not correlate well with each other, as was previously noted, they do have approximately the same coefficient with some sign. In both equations, this coefficient is the smaller, indicating this is the least important vaiable in describing K_{ow}. One can see from Table 2 that the t-score indicates that π, is not significant at the 95% level ($t_{.95}$=1.960).[24] The π, term has been retained for comparison with the LSER equation. The signs of the polarizability terms suggest that, because it is easier to polarize the electron cloud and induce a dipole, the solubility in water would increase.

The final descriptors, those dealing with the basicity, also yield a fair amount of information. First, the signs agree with each other and with chemical intuition. The degree of basicity should be proportional to the solubility in water and, therefore, inversely proportional to the K_{ow}. The β term does have a negative coefficient, indicating that an increase in basicity lowers the LOG K_{ow}. The TLSER parameters, εb and q_- produce a similar trend. As previously described, εb increases with a decrease in basicity. Therefore, the coefficient should be positive, indicating that a decrease in covalent basicity increases the solubility in octanol. The electrostatic term, q, is directly proportional to the basicity. Therefore, the coefficient should be negative as shown in equation 12. Equation 13 indicates β to be the single most important descriptor in the LSER. Table 2 shows that one of the basicity terms, q_-, is as significant as the molecular volume.

3.2.2   Konneman's Fish Toxicity.

A second type of solute/solvent interaction, one completely different from K_{ow}, is the application of the TLSER to toxicity and biological activity. The incidence of toxicity can be preceived in one of two mechanisms. One is by a specific receptor lock-and-key approach. The second is by non-specific general toxicity involving multiple-potential mechanisms. This section and the next one give two examples of the applicability of the TLSER to nonspecific toxicity.

16

Konneman measured the LC_{50} for a series of aliphatic and aromatic industrial pollutants on guppies (*Poecilia reticulata*).[25] Due to the diverse nature of the compounds measured, a single specific mechanism would not suffice. Equation 14 shows the TLSER for the LC_{50} of Konneman.[14]

$$\text{LOG } LC_{50} = -0.928 \left[\frac{V_{mc}}{100} \right] - 10.557 \, (\pi_I{*}10) - 1.442 \left[\frac{\epsilon_b}{10} \right] - 0.443 q_I + 18.082$$

$$ N = 32 \quad R = 0.9434 \quad SEE = 0.5736 $$

(14)

$$\text{LOG } LC_{50} = -3.157 \left[\frac{V_m}{100} \right] - 0.003 \pi^* + 3.277 \beta + 5.397$$

$$ N = 24 \quad R = 0.9778 \quad SEE = 0.3152 $$

(15)

As with the K_{ow}, the LSER yields a significantly (based upon the increased R and the decreased SEE) better fit. However, the coefficients in this case are quite different. The only possible explanation is that the additional compounds used for the TLSER were significantly different from the other compounds.

3.2.3 Microtox Toxicity Test.

A second biologically based interaction correlated for the effectiveness of the TLSER descriptors is the Microtox test.[26] The Microtox test is being used as a cost-effective prescreening technique for locating organic chemicals toxic to fish. In this test, the inhibition of bioluminescence in *Photobacterium phosphoreum* is used to determine the EC_{50} (50% education in light production after 5 min). Correlations of toxicant EC_{50} to *Photobacterium phosphoreum* and toxicant EC_{50} to the fathead minnow have shown fair to good results.

Equations 16 and 17 show the correlations derived from using the TLSER and the LSER parameters, respectively.[27*] The corresponding descriptors, in general, correlate well with each other in terms of sign and absolute magnitude. In this case, the correlations of the two parameter sets yield almost the same degree of fit based upon R and SEE.

$$\text{LOG } EC_{50} = -4.067 \left[\frac{V_{mc}}{100} \right] - 4.147 \, (\pi_I{*}10) + 3.902 q_- - 2.777 q H_+ + 11.656$$

$$ N = 25 \quad R = 0.9823 \quad SEE = 0.3323 $$

(16)

*Famini, G.R., Using Theoretical Descriptors in Structure Activity Relationships VI. A Comparison of QSAR Methodologies, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, unpublished data July 1989, UNCLASSIFIED Report.

$$\text{LOG EC}_{50} = -4.36\left[\frac{V_m}{100}\right] - 0.979\pi^* + 4.112\beta - 1.798\alpha + 7.531$$

(17)

$$N = 25 \quad R = 0.9889 \quad SEE = 0.2634$$

3.2.4 UV-Visible (UV-Vis) Absorption in Ylides.

The examples shown to this point have been solute effects. That is, the solvent environment has not changed, be it octanol/water, a micro-organism, or a fish, and the regressions explain the phenomena of a solute going into or through these systems. Another application of the LSER and the TLSER is the analogous relationship, understanding solvent effects. The solute remains unchanged, and the solvent system is altered, demonstrating the analogous relationship.

The LSER parameters, also called the solvatochromic parameters, were originally developed from UV-Vis spectral shift of selected indicators. As such, it would be expected that the LSER would yield excellent results. Molecular orbital (MO) calculations can predict neat, spectral abosrptions with varying degress of accuracy. However, MO calculations have not been capable of idenfitying and quanitifying solvent effects to spectral shifts in most cases.

To test the applicability of the TLSER parameters to solvent effects in general and UV-Vis absorption in particular, a data set of Dorohoi was used.[28] The solvent effect on the spectrum of the pyridinium ylide (Figure 4) was measured for 23 solvents.
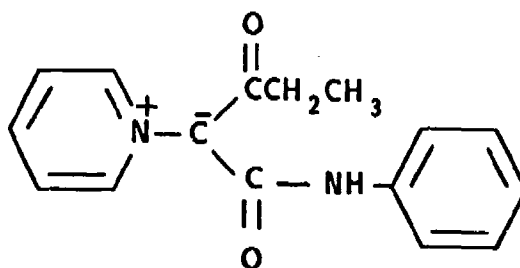

Figure 4. Pyridinium Ylide

The regressuion using the significant TLSER parameters is shown in equation 18. Equation 19 shows the analogous relationship for the LSER parameters.*

$$EA = 20545 - 2444.3 \left[\pi_I * 10 \right] + 3540 \left[\frac{\epsilon_a}{10} \right] + 1430q_- - 383qH_+$$

$$N = 23 \quad R = 0.9508 \quad SEE = 307.5 \tag{18}$$

$$EA = 20617 + 239\pi^* + 1357\beta + 1412\alpha$$

$$N = 23 \quad R = 0.9529 \quad SEE = 295.1 \tag{19}$$

These regressions show that although the LSER (equation 19) correlates better, the increase is relatively minor. Unlike the previous examples, several of the comparision are not valid. The basicity values, β and q_-, have approximately the same value and same sign. Based upon a positive ϵ_a and a negative qH_+, the TLSER acidity term would indicate a negative correlation with increasing acidity. The LSER, however, has a positive acidity term, α. In addition, although the TLSER and LSER have roughly the same magnitude for the polarizability terms, the signs are reversed.

3.2.5 High Performance Liquid Chromatography (HPLC) Retention Times (Hafkensheld Capacity Factors).

Another solute/solvent interaction that can be correlated with the LSER and TLSER parameters is chromatography. In either liquid or gas chromatography (GC), the solute establishes an equilibrium between the eluent and the solid phase. The retention times and capacity factors are dependent upon the established equibrium. As shown earlier, the TLSER and LSER parameters describe equilibrium conditions well, and therefore should correlate with many HPLC and GC indices.

An example to demonstrate this usefulness is the reversed-phase, HPLC retention index of Hafkensheid and Tomlinson.[29] The capacity factor of a 75% methanol-water mixture (k_{75}) was determined for 19 different solutes. Equation 20 shows the TLSER for this data set, and equation 21 shows the LSER for the set.[29]

$$LOG \; k_{75} = -0.083 + 1.185 \left[\frac{V_{mc}}{100} \right] - 0.610 \, (\pi_I * 10) - 2.290q_- - 1.077qH_+$$

$$N = 19 \quad R = 0.9859 \quad SEE = 0.0729 \tag{20}$$

$$LOG \; k_{75} = -0.552 + 0.952 \left[\frac{V_m}{100} \right] - 1.456\beta - 0.209\alpha$$

$$N = 21 \quad R = 0.9938 \quad SEE = 0.0492 \tag{21}$$

As can be seen, the LSER gives a better correlation for this data set. The coefficients for the two sets of parameters result in the same sign and approximately the same values. In both equations, the basic term is the single most important descriptor.

3.2.6   Hydrolysis of Organophosphorus Compounds.

A major solute/solvent interaction of interest to CRDEC is the rate constant for the hydrolysis of chemical nerve toxins. Hydrolysis of the toxin is important from two aspects. First, the ability of the decontamination materials applied to the toxin to neutralize it, and second, persistence of the toxin in the environment due to resistance to natural hydrolysis. As such, it is important to be able to predict the hydrolysis of potential toxins to assess their natural persistence. The mechanism for this reaction is given in Figure 5.[30,31]
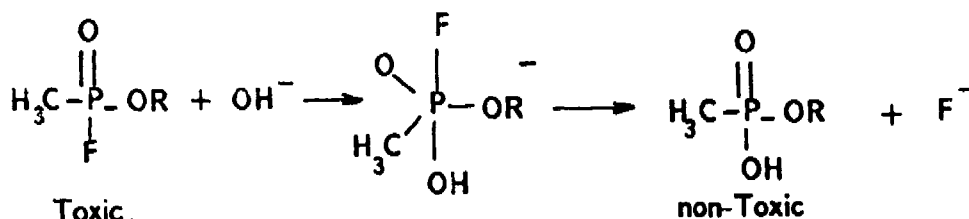
Figure 5.   Mechanism for the Base Hydrolysis of G-Agents

Equation 22 shows the TLSER regression foa a series of 10 organophosphorus compounds of the G-Agent type. No regression was possible for the Kamlet-Taft parameters due to the above empiricity of the solvatochromic parameters.

$$k_{OH} = -2814.4 - 97.7\left[\frac{V_{mc}}{100}\right] + 874.1\,(\pi_I * 10) + 1256\left[\frac{\varepsilon_b}{10}\right]$$

$$N = 10 \quad R = 0.9859 \quad SEE = 26.9$$

(22)

Several phenomena can be rationalized from equation 22. First, in examining the relative importance of each variable, it can be seen from Table 6 that the MO basicity is the most important parameter followed by V_m and π. At first glance, the mechanism would seem to indicate the solute, organophosphorus compound (OP), would be operating as the acid, and therefore ε_a should be the dominant term. This can be rationalized by remembering that, in general, as the E_{HOMO} decreases (i.e., ε_b increases), the E_{LUMO} decreases (i.e., the ε_a decreases). There is a direct but inverse relationship between ε_b and ε_a; as one increases, the other decreases.

Table 6.   t-Scores for Equation 22

Parameter	t-Score
V_{MC}	2.72
π_I	2.45
ε_b	11.66

The signs and magnitude of the Vm and π, can also be explained. As the OR group gets larger, it blocks the phosphorus from S_n2 attack by the OH^-. This will have a negative effect upon the rate contant, because in this data set, the principal variant is the ester organic group (OR). Therefore, V_{mc} will be directly proportional to the size of the OR group. In the TLSER, then, V_{mc} would be expected to be negative. On the other hand, the π, has a positive correlation with the rate constant. The OP will feel the effecive field due to the charge on the OH, and a dipole will be induced as the OH enters the reaction sphere of the OP. This induced dipole aids the hydrolysis and permits the base to react in a move facile manner.

4.   CONCLUSION

This report shows the general usefulness of the TLSER descriptors and the applicability to the correlation and prediction of solute/solvent interactions. This was a review of the previous reports in this series in which each descriptor has been presented individually. Figure 6 presents a summary of the types of solute/solvent interactions currently correlated with the TLSER parameters.

UV-Vis
Absorptions

Solubilities

TLSER

Parameters

Toxicities
and
Biological Activities
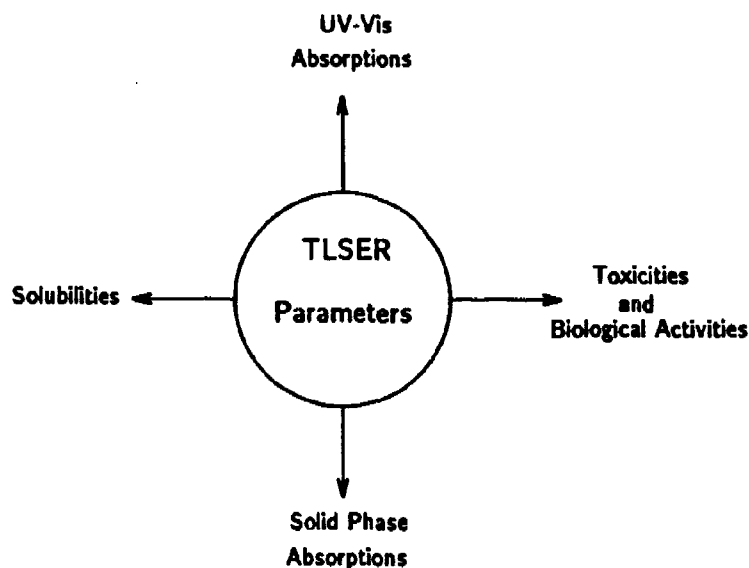
Solid Phase
Absorptions

Figure 6.   Correlations with the TLSER

In each of the correlations reported, the TLSER parameters yield a slightly lower correlation than the corresponding LSER parameters. However, it must be emphasized that the Kamlet-Taft LSER parameter uses several correction factors that increase the correlation coefficient and hence, the fit. The TLSER parameters use no such correction factors, and all values for the parameters are as they would appear in the calculations. The major drawback of the TLSER parameters is that intermolecular interactions are not taken into account. All calculations are done assuming a single molecule in

21

vacuum. This precludes the natural incorporation of crystal and proximity effects that are present in liquids and solids and in solution. The degree of fit of the TLSER should also increase as the level of the theoretical methods increase. Due to computational limitations, high level initial calculations are not feasible on compounds of this size at this time. However, as computational resources become more powerful, initial electronic results should further refine the TLSER.

The principal use of the TLSER is in the a priori prediction of solute/solvent interactions. Using any of the correlations in this study for new compounds requires only that the MNDO calculation be performed and the necessary volume calculation be done. Unlike the LSER parameters, synthesis is not required. As a result, the TLSER parameters for a given compound can usually be calculated.

LITERATURE CITED

1. Reid, R.C., Prausnitz, J.M., and Sherwood, T.K., The Properties of Gases and Liquids, 3rd ed, McGraw-Hill, New York, NY, 1979.

2. Lyman, W.J., Reehl, W.F., and Rosenblatt, D.H., Handbook of Chemical Property Estimation Methods, McGraw-Hill, New York, NY, 1981.

3. Famini, G.R., "Applications of Computational Chemistry to Simulant Technology," In Proceedings of the Annual Simulant Workshop, CRDEC-SP-002, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, December 1988, UNCLASSIFIED Report.

4. Hammett, L.P., "The Effect of Structure Upon the Reactions of Organic Compounds," J. Am. Chem. Soc. Vol. 59, p 96 (1937).

5. Hansch, C., "A Quantitative Approach to Biochemical Structure Activity Relationships," Acc. Chem. Res. Vol 2, p 232 (1969).

6. Shorter, J., "Linear Free Energy Relationships," Quarterly Review Vol. 27, p 44 (1970).

7. Kamlet, M.J., Abboud, J-L.M., Abraham, M.J., and Taft, R.W., "Linear Solvation Energy Relationships. 23. A Comprehensive Collection of the Solvatochromic Parameters, π, α, and β, and Some Methods for Simplifying the Generalized Solvatochromic Equation," J. Org. Chem. Vol. 48, pp 2877-2887 (1983).

8. Abboud, J-L.M., Kamlet, M.J., and Taft, R.W., "Journal Regarding a Generalized Scale of Solvent Polarities," J. Am. Chem. Soc. Vol. 91, pp 8325-8327 (1977).

9. Kamlet, M.J., and Taft, R.W., "Linear Free Energy Relationships. Local Empirical Rules or Fundamental Laws of Chemistry? A Reply to the Chemometricians," Acta Chem. Scand. Vol B39, pp 611-628 (1985).

10. Lewis, D.F.V., J. Comp. Chem. Vol 8, p 1084 (1987).

11. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships, I. Molecular Volume, CRDEC-TR-88031, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1988.

12. Wilson, L.Y., and Famini, G.R., Using Theoretical Descriptors in Structure Activity Relationships, III. Electronic Descriptors, CRDEC-CR-88083, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, August 1988, UNCLASSIFIED Report.

13. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships, II. Polarizability Index, CRDEC-TR-88137, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, September 1988, UNCLASSIFIED Report.

14. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships, IV. Molecular Orbital Basicity and Electrostatic Basicity, CRDEC-TR-013, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, November 1988, UNCLASSIFIED Report.

15. Dewar, M.J.K., and Thiel, W., "Ground States of Molecules, 38. The MNDO Method," J. Am. Chem. Soc. Vol 99, p 4899 (1977).

16. Stewart, J.J.P., MOPAC: A General Molecular Orbital Package, FJSRL-TR-86-0003, Frank J. Seiler Research Lab, U.S. Air Force Academy, CO, June 1986.

17. Hopfinger, A.J., "A Quantitative Structure Activity Relationship Investigation of Dehydrolfate Reductase Using Molecular Shape Analysis," J. Am. Chem. Soc. Vol 102, p 7126 (1980).

18. Leonard, J.M., A User's Guide to the Molecular Modeling Analysis and Display System, CRDEC-TR-030, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1989, UNCLASSIFIED Report.

19. Gupta, S., "QSAR Studies on Enzyme Inhibitors," Chem. Rev. Vol 87, p 1183 (1987).

20. Stewart, J.J.P., and Dewar, M.J.S., "A New Procedure for Calculating Molecular Polarizabilities Using MNDO," Chem. Phys. Lett. Vol. III, p 416 (1984).

21. Purcell, K.F., and Kotz, J.C., Inorganic Chemistry, p 531, W.B. Saunders Company, Philadelphia, PA, 1977.

22. Kamlet, M.J., Taft, R.W., Abboud, J-L.M., Famini, G.R., Doherty, R.M., and Abraham, M.J., "Solubility Parameters in Biological Media 5. An Analysis of the Physicochemical Properties Which Influence Octanol-Water Partition Coefficient," J. Pharm. Sci. Vol 4(8), p 807 (1985).

23. Hansch, C., and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley-Interscience, New York, NY, 1979.

24. Mendenhall, W., and Scheaffer, R.L., Mathematical Statistics and Applications, p 283, Doxbury Press, MA, 1973.

25. Konnemann, H., "Quantitative Structure Activity Relationships in Fish Toxicity Studies," Toxicology Vol. 19, p 209 (1981).

26. De Zwart, D., and Sloof, W., Aquat. Toxicol. Vol 4, p 129 (1983).

27. Kamlet, M.J., Taft, R.W., Vieth, G.D., Doherty, R.M., and Abraham, M.J., "Solubility Properties in Polymers and Biological Media, 7. An Analysis of Toxicant Properties That Influence Inhibition in Photobacterium Phosphoreum," Environmental Science and Technology Vol. 20, p 690 (1986).

28. Dorohoi, D., and Iancu, D., "The Influence of the Solvent on the Electronic Absorption Spectra of Some Pyridinium Ylides," Anale Stiintiche ale Universitatii Vol. 20, p 59 (1981).

29. Hafkensheid, T.L., and Tomlinson, E., "Isocratic Chromatographic Retention Data for Estimation Aqueous Solubilities," I.J. of Pharmaceutics Vol. 17, p 1 (1983).

30. Larsson, L., "The Alkaline Hydrolysis of Sarin," Acta Chem. Scan. Vol. II, p 1131 (1953).

30. Britton, K.B., Low Temperature Effects on Sorption, Hydrolysis and Photolysis of Organophosphates, Special Report 86-38, U.S. Army Corps of Engineers, Washington, DC, December 1986, UNCLASSIFIED Report.